BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

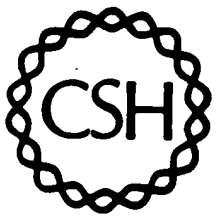
IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

Molecular Cloning

A LABORATORY MANUAL

T. Maniatis Harvard University
E. F. Fritsch Michigan State University
J. Sambrook Cold Spring Harbor Laboratory



Cold Spring Harbor Laboratory
1982

Molecular Cloning

A LABORATORY MANUAL

Printed in the United States of America
Book and cover design by Emily Harste

Front cover: The electron micrograph of bacteriophage λ particles stained with uranyl acetate was digitized and assigned false color by computer. *Thomas R. Broker, Louise T. Chow, and James I. Garrels*

Back cover: E. coli DH1 with fimbriae was negatively stained with phosphotungstic acid and the electron micrograph was digitized and assigned false color by computer. *Jeffrey A. Engler, Thomas R. Broker, and James I. Garrels*

Cataloging in Publications data

Maniatis, T.
Molecular cloning.

(A laboratory manual)
Bibliography: p.
Includes index.
1. Molecular cloning. 2. Eukaryotic cells. I. Fritsch, Edward F. II. Sambrook, Joseph. III. Title. IV. Series.
QH442.2.F74 574.87'3224 81-68891
ISBN 0-87969-136-0 AACR2

SAN 203-6185

7

Synthesis and Cloning of cDNA

The enzymatic conversion of poly(A)$^+$ mRNA to double-stranded cDNA and the insertion of this DNA into bacterial plasmids has become a fundamental tool of eukaryotic molecular biology (for reviews, see Efstratiadis and Villa-Komaroff 1979; Williams 1981). Although a number of different approaches to synthesizing double-stranded DNA copies of mRNA have been reported, the most commonly used procedure involves synthesis of the first cDNA strand with reverse transcriptase (RNA-dependent DNA polymerase), removal of the RNA template by alkaline degradation, synthesis of the second DNA strand with *E. coli* DNA polymerase I or reverse transcriptase (using a hairpin loop at the 3′ end of the first DNA strand as primer), and finally, digestion of the loop connecting the first and second cDNA strands with the single-strand-specific nuclease S1. In this chapter, we summarize important technical points for each step of the synthesis of double-stranded cDNA, and we then discuss procedures necessary for joining this DNA to plasmid cloning vectors.

Synthesis of cDNA

SYNTHESIS OF THE FIRST cDNA STRAND

A number of papers describing optimization of conditions for producing "full-length" cDNA transcripts have been published (Efstratiadis et al. 1976; Buell et al. 1978; Retzel et al. 1980). Different mRNAs are copied into DNA with different efficiencies; thus conditions that are optimal for copying one species of mRNA may not work as well for another. In general, when dealing with heterogeneous populations of mRNA, conditions are used that lead to the greatest overall yield of cDNA. The following parameters are important.

Reverse Transcriptase

The most important factor in the synthesis of long cDNAs is the quality of the reverse transcriptase used in the reaction. Until recently, the major producer of reverse transcriptase was Dr. J. W. Beard (Life Sciences, Inc., 1509½ 49th Street South, St. Petersburg, FL 33707), who provided the enzyme on contract to the National Institutes of Health. After the NIH program was terminated, Life Sciences, Inc., began selling the enzyme directly. Reverse transcriptase is also available commercially from Bethesda Research Laboratories (Gaithersburg, MD) and Boehringer Mannheim Biochemicals (Indianapolis, IN).

Although the quality of these enzymes is generally good, the amount of contaminating RNase varies from batch to batch. (Some suppliers assay for and provide information about contaminating RNase.) This problem can be circumvented by additional purification of the enzyme (Marcus et al. 1974; Faras and Dibble 1975; Kacian 1977; Myers et al. 1980) or by including potent inhibitors of RNase, such as vanadyl-ribonucleoside complexes or RNasin, in the reverse transcription reaction. Many factors previously thought to be important for efficient synthesis of full-length cDNA transcripts actually work by protecting the RNA template from RNases (Buell et al. 1978; Retzel et al. 1980). For example, the addition of sodium pyrophosphate or ribonucleoside triphosphates was originally thought to increase the efficiency with which reverse transcriptase copied RNA (Kacian et al. 1972). However, with highly purified reverse transcriptase, the addition of these compounds has no effect (Retzel et al. 1980).

The ratio of reverse transcriptase to mRNA template is also important in optimizing the yield of full-length cDNA (Friedman and Rosbash 1977). With a given amount of template, the yield and the size of the cDNA transcript increases with increasing amounts of reverse transcriptase. In one study, maximum yield of full-length transcripts was reached at 80 units of

enzyme per microgram of template, a 30-fold to 60-fold molar excess of enzyme to template (Friedman and Rosbash 1977). Such a high ratio of enzyme to template requires the use of highly purified enzyme and the inclusion of inhibitors of RNase in the reaction.

pH

A pH of 8.3 is optimal for efficient incorporation and production of full-length transcripts. A deviation of ± 0.5 pH units will result in a 5-fold decrease in the production of full-length transcripts. A number of buffer systems have been tested but none are better than Tris.

Monovalent Cation

Ionic conditions substantially affect the transcriptional efficiency of various templates. Longer transcripts are obtained with potassium than with sodium ions. The optimum potassium-ion concentration for both total synthesis and length of cDNA is 140–150 mM.

Divalent Cation

Divalent cations are an absolute requirement for reverse transcriptase activity. No activity is observed below 4 mM Mg^{++}; the optimum concentration for the production of full-length transcripts is 6–10 mM.

Deoxynucleoside Triphosphates

The use of high concentrations of each of the four deoxynucleoside triphosphates (dNTPs) is particularly important for efficient cDNA synthesis (Efstratiadis et al. 1976; Retzel et al. 1980). If the concentration of only one of them drops below 10–50 μM, the yield of full-length transcripts decreases significantly. Using avian myeloblastosis virus (AMV) RNA as a template, maximum production of full-length cDNAs was achieved at a concentration of 75 μM of all four dNTPs (Retzel et al. 1980). However, since little or no inhibition of transcription is observed in the 100 μM to 1 mM range, dNTP concentrations of 200–250 μM are generally used.

SYNTHESIS OF THE SECOND cDNA STRAND

For reasons that are not yet understood, the 3′ ends of single-stranded cDNAs are capable of forming hairpin structures and therefore can be used to prime the synthesis of the second cDNA strand by *E. coli* DNA polymerase I or reverse transcriptase (see Fig. 7.1). Although there has been a great deal of speculation regarding the structure of the hairpin loops at the end of cDNAs and the mechanism by which they are generated, the phenomenon has not been systematically studied.

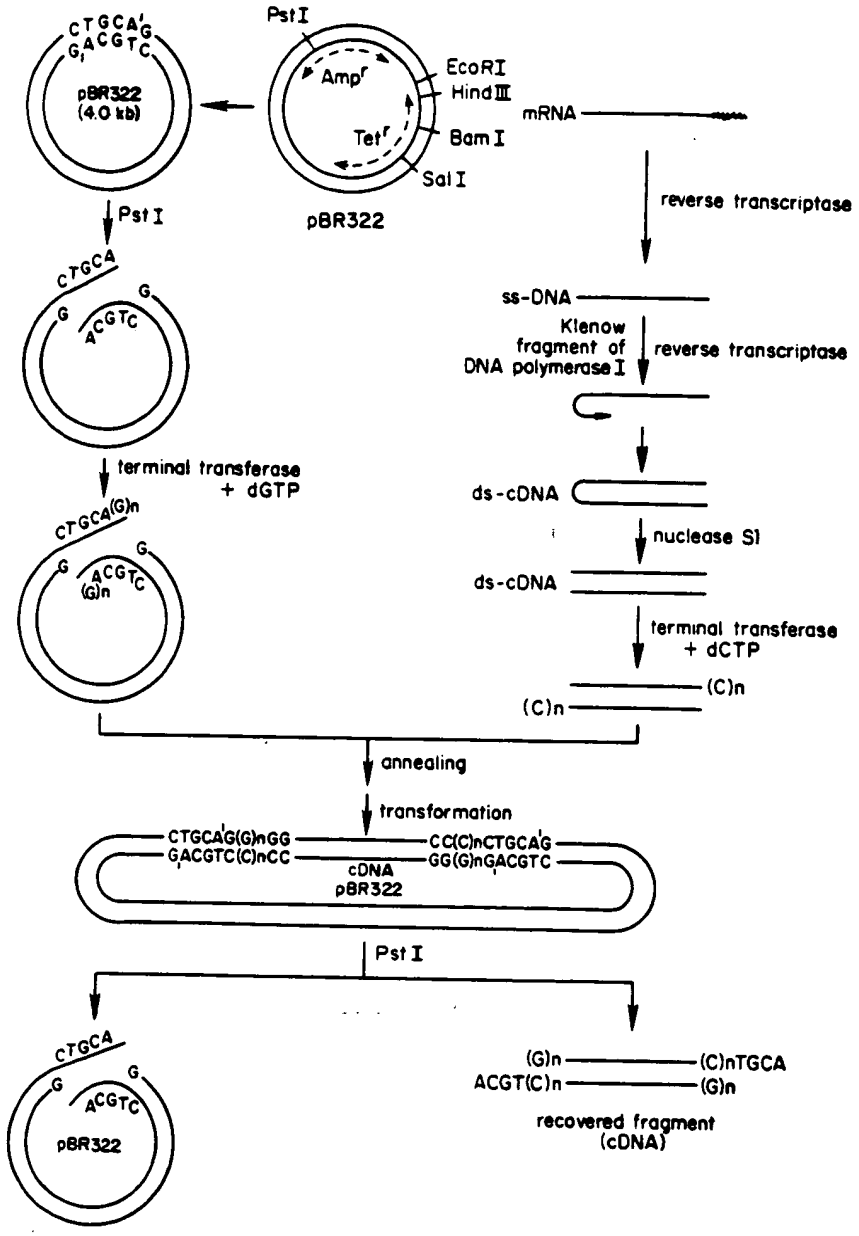


Figure 7.1

The conditions first used to achieve full-length, second-strand cDNA synthesis by DNA polymerase I (Efstratiadis et al. 1976) are still widely used (Wickens et al. 1978). In brief, the reaction is carried out at pH 6.9 to minimize the 5′ → 3′ exonuclease activity of DNA polymerase I and at 15°C to minimize the possibility of synthesizing "snapback" DNA. The Klenow fragment of DNA polymerase I, which lacks the 5′ → 3′ exonuclease activity, has also been successfully employed to synthesize the second cDNA strand.

Many investigators have utilized reverse transcriptase to synthesize the second cDNA strand using conditions similar to those already described for the first cDNA strand synthesis. Although there is one report that AMV reverse transcriptase could not be used to synthesize the second strand of an immunoglobulin cDNA (Rougeon and Mach 1976), the success of a large number of experiments that used reverse transcriptase for second-strand synthesis indicates that this is not a general problem. We recommend using both enzymes in succession. The rationale of this procedure, which was suggested by A. Efstratiadis, is that DNA polymerase I and reverse transcriptase may pause or stop at different sequences. Thus, partially synthesized second strands produced by one enzyme may be extended to completion by the other.

CLEAVAGE OF THE HAIRPIN LOOP WITH NUCLEASE S1

After synthesis of cDNA is complete, the first and second strands are covalently joined by the hairpin loop that was used to prime the second-strand synthesis (Efstratiadis et al. 1976). This loop is susceptible to cleavage by the single-strand-specific nuclease S1. The resulting termini are not always perfectly blunt-ended, and the efficiency of cloning is improved if they are repaired with the Klenow fragment of *E. coli* DNA polymerase I (Seeburg et al. 1977). The duplex DNA is then either fractionated according to size and the largest molecules inserted into bacterial plasmids, or an entire spectrum of sizes of double-stranded DNA is cloned to generate a cDNA library.

Molecular Cloning of Double-stranded cDNA

A variety of methods has been used to link double-stranded cDNA to plasmid vectors (Efstratiadis and Villa-Komaroff 1979; Maniatis 1980; Williams 1981). The most commonly used procedures are:

1. The addition of complementary homopolymer tracts to double-stranded cDNA and to the plasmid DNA. The vector and double-stranded cDNA are then joined by hydrogen bonding between the complementary homopolymeric tails to form open circular, hybrid molecules capable of transforming *E. coli*. The formation of closed circular DNA by in vitro enzymatic ligation is not necessary to establish the recombinant plasmids in *E. coli*.

2. The addition of synthetic linkers to the termini of double-stranded cDNA. After cleavage with the appropriate restriction enzyme, the cDNA molecules are inserted into plasmid DNA that has been cleaved with a compatible enzyme.

HOMOPOLYMERIC TAILING

dA · dT Tailing

Calf-thymus terminal deoxynucleotidyl transferase, which catalyzes the addition of deoxynucleotides to the 3′-hydroxyl ends of single- or double-stranded DNA, was first used by Wensink et al. (1974) to introduce recombinant DNA into *E. coli* by a dA · dT joining procedure (Jackson et al. 1972; Lobban and Kaiser 1973). In the original procedure, a small number of nucleotides were removed from the 5′ ends of the duplex DNAs to leave protruding, single-stranded, 3′-hydroxyl termini, which served as efficient templates for terminal transferase. The need for this step was obviated when it was shown that terminal transferase could utilize recessed 3′ termini in the presence of cobalt ions (Roychoudhury et al. 1976). Usually, 50 to 150 dA residues are added to the linearized vector DNA and a corresponding number of dT residues to the double-stranded cDNA.

Double-stranded cDNA inserted into plasmids via the dA · dT joining procedure can be excised and recovered in one of three ways. The first method involves digestion of the recombinant plasmid DNA with nuclease S1 under moderately denaturing conditions, which cause preferential melting of the dA · dT linkers (Hofstetter et al. 1976). This is achieved by including formamide in the digestion buffer (25–50%) and carrying out the digestion at an elevated temperature (37–55°C). The efficiency of this reac-

tion depends on the length of the dA · dT linkers: Inserts with short linkers are difficult to excise. The optimal conditions (including DNA and enzyme concentrations) should be determined empirically for each cDNA clone. In some cases, higher yields of the insert and fewer extraneous cleavage products are obtained when the single-strand-specific, mung-bean nuclease is used rather than nuclease S1 (M. R. Green, unpubl.). This may be related to the relatively high specificity of mung-bean nuclease for AT-rich duplex DNA (Johnson and Laskowski 1970).

An alternative but seldom-used procedure involves the conversion of plasmid DNA to linear duplex molecules using a restriction enzyme that does not cleave within the inserted DNA (Goff and Berg 1978). The linear DNA is then denatured and briefly renatured to allow snapback structures to form between the dA and dT residues that flank the cDNA insert on each strand. The vector DNA is then removed by treatment with *E. coli* exonuclease VII, which digests single-stranded DNA in both the 5′ → 3′ and 3′ → 5′ directions. The dA · dT duplex is then melted, and the two strands of the insert are reannealed to form duplex DNA.

Another strategy is to insert the double-stranded cDNA into a site that is closely flanked by two hexanucleotide restriction sites: In pBR322, for example, the sites for *Eco*RI, *Cla*I, and *Hin*dIII occur within a 30-bp region. Thus, by inserting the double-stranded cDNA into the *Cla*I site via dA · dT tailing, the insert can be recovered by digesting with *Eco*RI and *Hin*dIII.

In principle, *Hin*dIII sites can be regenerated by digesting plasmid DNA with *Hin*dIII, tailing with oligo(dT), and annealing with dA-tailed, double-stranded cDNA. The cDNA insert can then be recovered by *Hin*dIII digestion. For reasons that are not clear, this method has been used only rarely.

dC · dG Tailing

Currently, the most widely used procedure for cloning cDNAs by homopolymeric tailing involves addition of dG tails to the plasmid and complementary dC tails to the cDNA (Villa-Komaroff et al. 1978; Rowekamp and Firtel 1980). This method yields clones from which inserts can be easily removed. A plasmid such as pBR322 or pAT153 is digested with the enzyme *Pst*I, which cleaves the sequence

```
5'
  C-T-G-C-A-G 3'
3' G-A-C-G-T-C
                5'
```

leaving protruding 3′ tails. The addition of a short stretch of dG residues to the linear plasmid DNA results in regeneration of a *Pst*I site at each end of the insert, which can therefore be recovered from the plasmid by digestion with *Pst*I (see Fig. 7.1). In practice, the efficiency of regenerating the *Pst*I site depends on the quality of *Pst*I used to linearize pBR322 DNA and the quality of the terminal transferase. If the penultimate residue is removed

from the protruding 3′ tail by trace amounts of exonuclease, subsequent addition of dG residues will not recreate a *PstI* recognition sequence. Given reasonable care, however, as high as 80–90% of the recombinant plasmids constructed by this method contain inserts flanked by *PstI* sites.

Recently, the number of dA · dT and dG · dC residues required for optimal efficiencies of DNA transformation was determined (Peacock et al. 1981). In general, the number of residues on the plasmid and the cDNA should be approximately equal, with approximately 100 residues being added to each DNA for dA · dT joining and approximately 20 for dG · dC joining (Peacock et al. 1981). Interestingly, the bacterial strain can make a significant difference to the transformation efficiency. RR1, a *recA*$^+$ strain of *E. coli*, yielded 10 times as many recombinant cDNA clones made by the dA · dT tailing procedure as did the *recA*$^-$ host HB101. In the same experiment, untreated pBR322 DNA transformed the two strains with equal efficiency (Peacock et al. 1981). It would therefore appear that the bacterial *recA* system is involved in repairing open circular, hybrid DNA molecules that contain homopolymer tails.

SYNTHETIC DNA LINKERS

Synthetic linkers containing one or more restriction sites provide an alternative method to join double-stranded cDNA to plasmid vectors. Double-stranded cDNA, generated as described earlier, is treated with bacteriophage T4 DNA polymerase or *E. coli* DNA polymerase I, enzymes that remove protruding, 3′, single-stranded termini with their 3′ → 5′ exonucleolytic activities and fill in recessed 3′ ends with their polymerizing activities. The combination of these activities therefore generates blunt-ended cDNA molecules, which are then incubated with a large molar excess of linker molecules in the presence of bacteriophage T4 DNA ligase, an enzyme that is able to catalyze the ligation of blunt-ended DNA molecules. Thus, the products of the reaction are cDNA molecules carrying polymeric linker sequences at their ends. These molecules are then cleaved with the appropriate restriction enzyme and ligated to a plasmid vector that has been cleaved with a compatible enzyme.

The double-stranded cDNA molecules containing the synthetic cohesive ends will, of course, ligate to each other as well as to the vector DNA. In addition, the vector can recircularize by self-ligation and increase the background of nonrecombinant plasmids. These problems can be circumvented to a large extent by treating the linearized plasmid with phosphatase (see page 133) and/or by ligating different linkers to each end of the cDNA. In the original description of this method (Kurtz and Nicodemus 1981), two different linkers were simultaneously ligated to cDNA. During this process, it would be expected that 50% of the cDNA molecules would receive the same linker at each end; such molecules could not be inserted into plasmid DNA by directional cloning. In practice, this figure is even greater because one linker almost always has a higher rate of ligation to cDNA than the other.

This problem can be solved by adding one linker to the cDNA before cleaving the hairpin loop with nuclease S1 and the second linker after the S1 treatment. The double-linkered cDNA can then be treated with the appropriate restriction enzymes and inserted into a plasmid vector by directional cloning (see Fig. 7.2).

This opens the possibility of inserting cDNA in the correct orientation into vectors that allow expression of the inserted sequences in bacteria (see Chapter 12) and of identifying clones of interest by screening bacterial colonies for the presence of material that reacts with specific antisera to a particular gene product. This technique could be of great value when cloning rare mRNAs, for which no nucleic acid probes are available.

One problem with this approach is that the double-stranded, cDNA linker-DNA hybrids must be digested with the appropriate restriction enzymes to generate cohesive ends (Scheller et al. 1977). If the double-stranded cDNA contains one or more recognition sites for either one of the enzymes, it will be cleaved and subsequently cloned as two or more DNA fragments, making the structural analysis of the full-length cDNA difficult. This problem can be alleviated by using synthetic linkers carrying recognition sequences for restriction enzymes that cleave mammalian DNA very rarely (e.g., *Sal*I), by using *Eco*RI methylase to protect the DNA from cleavage with *Eco*RI, or by using synthetic adapters rather than linkers. Adapters are short, synthetic, double-stranded cDNAs that are blunt at one end and cohesive at the other (e.g., a *Hin*dIII cohesive end). By placing a 5′ phosphate on the blunt end of the adapter and a 3′ hydroxyl on the sticky end, the adapter will ligate to blunt-ended, double-stranded cDNA but not to itself. Unlike linkers, adapters do not have to be digested with restriction enzymes prior to ligation to double-stranded cDNA.

OTHER METHODS OF CLONING cDNA

Most of the cDNA clones thus far characterized have been constructed by using one of the techniques described above. Below we briefly describe three alternative procedures for cDNA cloning. The first procedure, mRNA · cDNA hybrid cloning, has limited applicability because of its low efficiency. The second procedure involves second-strand cDNA synthesis primed by oligonucleotides, while the third method involves plasmid-primed, first- and second-strand cDNA synthesis. Although the latter two procedures have not yet been widely applied and we ourselves have no direct experience with them, the published reports indicate that both provide an efficient means of obtaining full-length cDNA clones.

mRNA · cDNA Cloning

Another method for cDNA cloning involves transformation of *E. coli* with mRNA · cDNA hybrids that have been joined to plasmid vectors (Wood and Lee 1976; Zain et al. 1979). The bacterial host removes the mRNA and replaces it with DNA. After the first strand of cDNA has been synthesized in

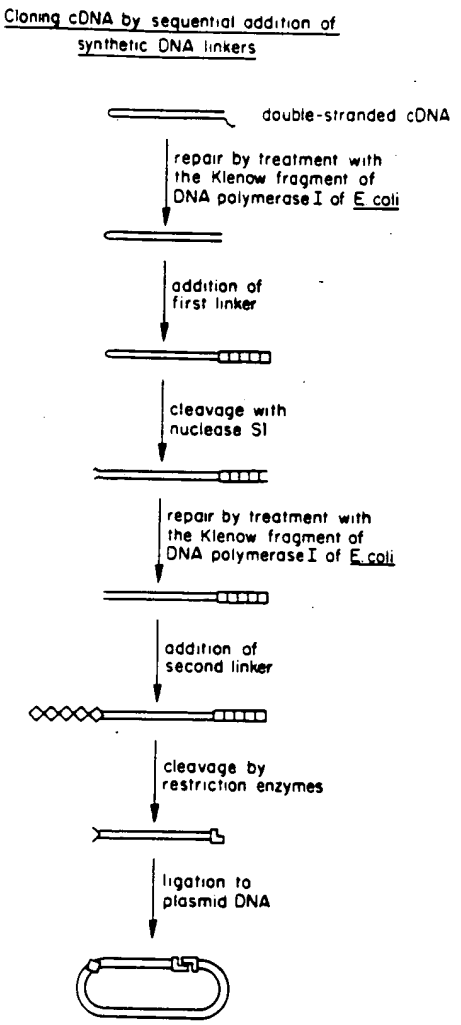


Figure 7.2

Cloning cDNA by sequential addition of synthetic DNA linkers.

the usual way, dA residues are added to the mRNA · cDNA hybrid, and the tailed hybrid is then annealed to a plasmid tailed with dT. Because the efficiency of tailing the 3′-hydroxyl group of RNA is at least 10 times less than the homologous reaction with DNA, most of the dA residues added to the hybrid are incorporated at the 3′ end of the DNA strand. Joining of the other end of the hybrid to the vector is probably accomplished by hydrogen bonding between the tract of natural poly(A) at the 3′ end of the mRNA and the dT-tailed plasmid. The two practical advantages of this procedure are (1) that no synthesis of second cDNA strand is required and (2) that cleavage of the DNA hairpin by nuclease S1 is not necessary. Furthermore, the procedure should in theory allow the sequences at the 5′ end of the mRNA (which are normally lost during nuclease-S1 cleavage) to be cloned. Its major disadvantage, however, is that it is at least 10 times less efficient than double-stranded cDNA cloning and is therefore unsuitable for constructing large numbers of cDNA clones.

Second-strand cDNA Synthesis Primed by Oligonucleotides

Synthesis of the second strand of cDNA is usually primed by hairpin structures at the 3′ terminus of the first strand. An alternative procedure is to tail the first strand of cDNA directly with dT (Rougeon et al. 1975) or dC (Land et al. 1981). The second strand is then synthesized using an oligo(dA) or oligo(dG) primer, respectively, producing duplex cDNA flanked by duplex homopolymeric tracts at each end. The duplex DNA is then tailed with dC and inserted into a plasmid that has been cleaved with *Pst*I and tailed with dG.

The chief advantage of this procedure is that it eliminates the difficult step in which nuclease S1 is used to cleave the hairpin loop in double-stranded cDNA and thus facilitates the efficient cloning of full-length, double-stranded cDNA. One potential pitfall in this procedure is that even highly purified preparations of terminal transferase are contaminated with single-strand-specific nucleases. Presumably, this latter problem could be circumvented by tailing the first cDNA strand as a DNA · RNA hybrid.

Plasmid-primed, First- and Second-strand cDNA Synthesis

Recently, a novel method for high-efficiency cloning of full-length, double-stranded cDNA was published by Okayama and Berg (1982). The steps in their protocol are as follows (see Fig. 7.3A,B,C):

1. A plasmid primer for cDNA synthesis is prepared by dT tailing with terminal transferase. A fragment containing one of the dT tails, the bacterial origin of replication, and the ampicillin-resistance gene is prepared by digestion with a second enzyme, followed by agarose gel electrophoresis and oligo(dA) cellulose chromatography (Fig. 7.3A).

2. An oligo(dG)-tailed linker DNA is prepared by dG tailing a *Pst*I DNA fragment with terminal transferase, followed by digestion with a second enzyme to separate the two ends. The desired end fragment is purified by agarose gel electrophoresis (Fig. 7.3B).

3. The dT-tailed vector-primer is annealed with poly(A) mRNA at a molar ratio of 1.5–3 (mRNA:vector-primer), and a first cDNA strand is synthesized with reverse transcriptase (Fig. 7.3C).

4. dC tails are added to the 3′ end of the cDNA copy while it is still hydrogen bonded to the mRNA template. The dC tail added at the other end of the vector is then removed by restriction endonuclease digestion.

5. The oligo(dG)-tailed cDNA · mRNA plasmid is annealed and ligated to the oligo(dG)-tailed linker DNA.

6. The mRNA strand is replaced by DNA using the combined activities of RNase H, which degrades the RNA strand in an RNA · DNA hybrid, *E. coli* DNA polymerase I, which carries out a nick-translation repair of the second cDNA strand, and DNA ligase, which covalently closes the circular DNA molecule.

Okayama and Berg find that full-length or nearly full length cDNA copies are preferentially converted to duplex cDNA, and an efficiency of approximately 100,000 transformants per microgram of starting mRNA is obtained. The preferential cloning of long cDNA transcripts is thought to be a consequence of the preferential utilization of full-length reverse transcription by terminal transferase. They speculate that shortened or truncated cDNA strands in the mRNA · DNA duplex are not efficiently recognized by the terminal transferase and are therefore selected against. Although the rabbit α- and β-globin mRNA was used to establish this cDNA cloning procedure, Okayama and Berg indicate that other cDNA clones representing both rare and long (6500-nucleotide) mRNAs have been obtained with this procedure.

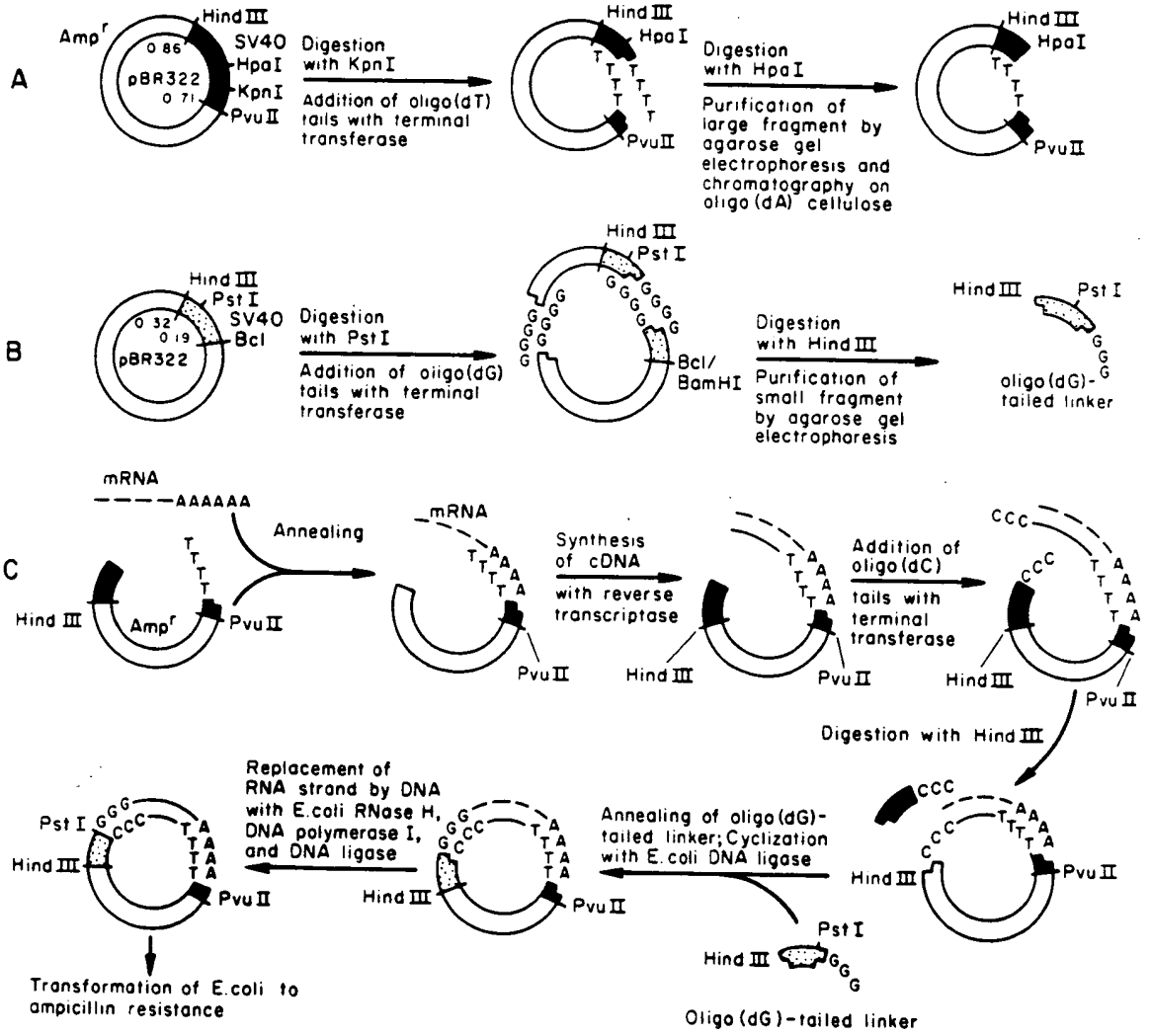


Figure 7.3

Preparation of (*A*) plasmid primer and (*B*) oligo(dG)-tailed linker DNA. (*C*) Steps in the construction of plasmid-cDNA recombinants. pBR322 DNA is represented by the open sections of each ring; SV40 DNA is indicated by the darkened or stippled segments. The numbers next to the restriction site designations are the corresponding SV40 DNA map coordinates.

Strategies for cDNA Cloning

ABUNDANT mRNAs

Initially, cDNA cloning was used to obtain copies of abundant mRNAs such as globin and ovalbumin. In these cases, the RNA of interest comprises as much as 50–90% of the total poly(A)$^+$ cytoplasmic RNA isolated from certain specific cell types. Consequently, no further purification of the particular mRNA is required before double-stranded cDNA is synthesized and cloned.

To identify cDNA clones of abundant mRNAs, transformed bacteria are assayed by nucleic acid hybridization for the presence of the desired DNA sequences. The probes consist either of ^{32}P-labeled, single-stranded cDNA synthesized in vitro by reverse transcriptase, using as template mRNA preparations that are rich in the sequences of interest, or of a partially fragmented, end-labeled preparation of the mRNA itself. As a good approximation, the mRNA sequences of interest will be represented both in the cloned, double-stranded cDNAs and in the probe in proportion to their abundances in the starting population. In cases like ovalbumin and globin, the chances are high that any colony hybridizing strongly to the probe will contain the desired DNA sequences.

Proof of the identity of the clone can be obtained in one of three ways:

1. By showing that the cloned cDNA is able to select the mRNA of interest from the starting population of mRNA. Usually the cloned cDNA is immobilized on a nitrocellulose filter and hybridized to mRNA in solution. After extensive washing, the mRNA is released from the hybrid and translated in a cell-free, protein-synthesizing system (hybridization/selection) (Goldberg et al. 1979).

2. By showing that the cloned cDNA is able to hybridize to the mRNA of interest and thereby inhibit its translation in vitro (hybrid-arrested translation) (Paterson et al. 1977).

3. By direct DNA sequencing. When the amino acid sequence of the protein product is known, it is a simple matter to establish that the cloned cDNA and the protein are colinear. Rapid methods have recently been developed to apply the Maxam-Gilbert (1977) or the Maat-Smith (1978) techniques to obtain the sequence of DNA fragments cloned in plasmids (Frischauf et al. 1980).

LOW-ABUNDANCE mRNAs

With refinements of methods for the efficient introduction of recombinant cDNA plasmids into *E. coli* (Hanahan and Meselson 1980) and for screening large numbers of transformed bacterial colonies for foreign DNA sequences (D. Hanahan, unpubl.), the cloning of mRNAs of relatively low abundance is possible.

The strategy currently employed involves the construction of large numbers of cDNA clones from total poly(A)$^+$ mRNA and the identification of the cDNA clones of interest. The entire collection of cDNA clones from a particular preparation of poly(A)$^+$ RNA is called a cDNA library.

A typical mammalian cell contains between 10,000 and 30,000 different mRNA sequences (Davidson 1976). Williams (1981) has determined the number of clones necessary to obtain a complete cDNA library from a human fibroblast cell that contains approximately 12,000 different mRNA sequences. The low-abundance class of mRNAs (< 14 copies/cell) comprises approximately 30% of the mRNA, and there are about 11,000 different mRNAs in this class. The minimum number of clones required to obtain a complete representation of low-abundance mRNA sequences is therefore 11,000/.30 = ~37,000. Of course, because of sampling variation and of preferential cloning of certain sequences, a much larger number of recombinants must be obtained to increase the chance that any given clone will be represented in the library. The number of clones required to achieve a given probability that any given low-abundance sequence will be present in a library is:

$$N = \frac{\ln (1 - P)}{\ln (1 - 1/n)}$$

where

N = number of clones required;
P = probability desired (usually 0.99);
n = fractional proportion of the total mRNA population that a single type of low-abundance mRNA represents.

Therefore, to achieve a 99% probability of obtaining a particular low-abundance mRNA from the human fibroblast described above:

P = 0.99
n = 1/37,000
N = 170,000

This number is within reach of existing techniques, since between 1×10^5 and 6×10^5 colonies per μg of double-stranded cDNA can be obtained either by homopolymeric tailing or by double-linker procedures.

A major problem, however, is the detection of extremely low abundance mRNA sequences by in situ colony hybridization (Gergen et al. 1979; Willi-

ams and Lloyd 1979; Dworkin and Dawid 1980). Several authors have calculated that clones representing as little as 0.05%-0.1% of the total mRNA molecules can be detected when in vitro labeled mRNA (or cDNA) is used as a probe. In practice, however, with the concentrations of probe that are usually available and with hybridization and autoradiographic exposure times that are reasonable, it is extremely difficult to detect clones containing cDNA complementary to mRNA species that are present in the initial population at less than 1 part in 200.

So far, no general method has been developed to clone such molecules. However, there are several techniques that may be used singly or in combination to deal with the problems encountered in identifying cDNA clones of RNAs that are only minor components of the total population and for which no hybridization probes are available.

Size Fractionation

The simplest technique is to fractionate the mRNA by size, for example, by density gradient centrifugation or gel electrophoresis under denaturing contions (see pages 199–206). Each fraction of the mRNA is then translated in vitro and the protein product of interest is identified by a combination of immunoprecipitation and SDS-polyacrylamide gel electrophoresis. The degree of enrichment obviously varies from mRNA to mRNA, depending on its size relative to the bulk of the mRNA population. At best, an enrichment of perhaps 10-fold may be attained; however, this may be sufficient to bring the mRNA within cloning range.

An alternative strategy is to construct a cDNA library from a partially enriched mRNA population obtained, perhaps, by sucrose gradient centrifugation, and then to screen the library by hybridization to probes synthesized by reverse transcription of a still more highly enriched mRNA population obtained by fractionation of mRNA through density gradients *and* denaturing gels. The aim is to reduce the library to a manageable number of cDNA clones that can be screened individually or in small batches by hybrid selection.

Synthetic Oligodeoxynucleotides

Purification of an mRNA present in low concentrations can be arduous and difficult. If a partial or complete amino acid sequence of the protein of interest is available, the method of choice involves the chemical synthesis of oligonucleotides complementary to the mRNA. The sequence of such oligonucleotides can be deduced from favorable short sequences of amino acids (Wu 1972). In essence, one scans the known protein sequence for areas rich in amino acids specified either by a single codon (e.g., methionine, AUG; tryptophan, UGG) or by two codons (e.g., phenylalanine, UUU, UUC; tyrosine, UAU, UAC; histidine, CAU, CAC). Knowing the frequency with which different degenerate codons are used (e.g., glutamine is usually specified by CAG) and by taking advantage of G · T base-pairing, it is often possible to

narrow down the candidate oligonucleotides to a manageable number. These oligonucleotides are then synthesized in vitro by either the phosphodiester method (Agarwal et al. 1972) or the more-efficient phosphotriester method (Hsiung et al. 1979). When these oligonucleotides are incubated under carefully defined annealing conditions with total poly(A)$^+$ mRNA, they form hybrids only with those species of mRNA to which they are exactly complementary. They can therefore be used as primers in reverse transcription reactions with unfractionated poly(A)$^+$ RNA to synthesize single-stranded probes for screening cDNA libraries (Chan et al. 1979; Noyes et al. 1979; Goeddel et al. 1980a; Houghton et al. 1980). If the synthetic oligodeoxynucleotides are sufficiently long (14–20 nucleotides), they can be used directly as probes to screen cDNA libraries for the clones containing sequences of interest (Montgomery et al. 1978; Goeddel et al. 1980a; Suggs et al. 1981).

A useful approach is to synthesize chemically a mixture of oligonucleotides that represent all possible coding combinations for a small portion of the amino acid sequence of the protein of interest (Wallace et al. 1979, 1981). One of these oligonucleotides will form a perfectly base-paired duplex with the double-stranded DNA, whereas the other oligonucleotides will form mismatched duplexes. If hybridization conditions of the appropriate stringency are chosen, only the perfectly matched duplex will be stable. This approach was recently employed to isolate cloned cDNA sequences for human β_2-microglobulin (Suggs et al. 1981). Note that the conditions used to screen colonies by hybridization are considerably more stringent than the conditions used to anneal primers to mRNA. Thus, when oligonucleotides are used as probes, they are much more specific than when used to prime cDNA synthesis on an mRNA template.

Oligonucleotides complementary to the coding region of an mRNA can never prime the synthesis of full-length cDNA molecules in reverse transcription reactions. Such oligonucleotides are therefore hardly ever used as primers to synthesize cDNA for cloning purposes. Their outstanding virtue is that they are (or can be used to generate) highly specific probes. The sensitivity of screening cDNA libraries is thereby increased to the level where clones synthesized from extremely rare mRNAs can easily be detected.

Differential Hybridization

This method has been used when two mRNA preparations are available that contain many sequences in common but that are different from each other in the presence and absence of a few species of interest. Examples of such sibling pairs might be mRNAs extracted from cells before and after exposure to heat shock, drugs, or hormones. In the simplest application of this technique, ^{32}P-labeled cDNA is synthesized in vitro from both preparations of poly(A)$^+$ RNA. Most of the cDNA sequences will be shared by the two preparations. However, the cDNA synthesized from the induced-cell RNA should contain additional sequences complementary to any new species of poly(A)$^+$ RNA. The two probes are then used to screen replicas of a cDNA

library constructed from mRNA extracted from the induced-cell population. Those colonies hybridizing specifically to the induced-cell cDNA probe are likely to contain cloned copies of the induced mRNAs. Examples of inducible genes cloned in this way include the galactose-inducible genes of yeast (St. John and Davis 1979) and human fibroblast interferon (Taniguchi et al. 1980a). This procedure has also been used to identify cDNA clones of developmentally regulated mRNAs from *Xenopus laevis* (Dworkin and Dawid 1980), *Dictyostelium discoidum* (Williams and Lloyd 1979; Rowekamp and Firtel 1980), and sea urchins (Lasky et al. 1980).

cDNA clones corresponding to developmentally regulated mRNAs can also be identified using another type of differential hybridization. A population of cDNA molecules enriched in sequences characteristic for a particular developmental stage is used to probe a cDNA or genomic library (Timberlake 1980; Zimmerman 1980). This enrichment is accomplished by "cascade hybridization" in which cDNA prepared from mRNA obtained at one developmental stage (stage 1) is hybridized to a 20-fold excess of mRNA obtained from another stage (stage 2). The mRNA · cDNA hybrid is then removed by binding to hydroxyapatite. This procedure is repeated twice more using a 50-fold to 100-fold excess of stage-2 mRNA. The final, unbound cDNA fraction is then hybridized to a 100-fold excess of stage-1 mRNA, and the hybrid is recovered from hydroxyapatite. After removing the mRNA by alkaline hydrolysis, the cDNA that is highly enriched in stage-1-specific sequences is used to probe a stage-1 cDNA library.

Immunopurification of Polysomes

One approach to enriching specific mRNAs is to purify particular polysomes by virtue of the reaction between antibodies and nascent polypeptide chains (Cowie et al. 1961). The technique, which originally involved immunoprecipitation of polysomes, was limited to mRNAs that encode abundant proteins such as ovalbumin (Palacios et al. 1972) and immunoglobulin (Schechter 1973). Attempts to apply the method to mRNAs of lesser abundance were disappointing (Flick et al. 1978). However, recently the use of immunoaffinity columns (Schutz et al. 1977) and protein A–Sepharose columns (Shapiro and Young 1981) has resulted in significant improvements of the technique. For example, a relatively abundant trypanosome surface-antigen mRNA was purified by reacting polysomes with a heterogeneous antiserum to the surface antigen and trapping the complex on a protein A–Sepharose column (Shapiro and Young 1981). Lower-abundance mRNAs now can be isolated by combining the use of protein A–Sepharose with the use of monoclonal antibodies. Korman et al. (1982) used a monoclonal antibody to the heavy chain of the human HLA-DR antigen to purify the corresponding mRNA, which represents only 0.01–0.05% of the total mRNA. These investigators report a 2000-fold to 3000-fold purification of the HLA-DR mRNA. The purified mRNA can then be used to prepare a cDNA probe for screening a total cDNA library, or it can be used directly to prepare a double-stranded cDNA clone.

Procedures for cDNA Cloning

On the following pages, we describe in detail two methods for cDNA cloning using either dG · dC homopolymer tailing or the double-linker technique. We have used both methods successfully to produce cDNA libraries that appear to reflect the complexity of mRNA populations extracted from several types of mammalian cells.

The method for homopolymer tailing is a synthesis of protocols published by a number of different groups, in particular Efstratiadis and Villa-Komaroff (1979), Rowekamp and Firtel (1980), and B. Roberts (pers. comm.). The method utilizing sequential addition of linkers is an unpublished modification by J. Fiddes of a protocol devised by Kurtz and Nicodemus (1981).

SYNTHESIS OF DOUBLE-STRANDED cDNA

The conditions given below are optimal for synthesis of cDNA from heterogeneous populations of mRNA. However, individual species of mRNA may be copied by reverse transcriptase at different efficiencies.

First-strand Synthesis

1. Purify poly(A)$^+$ mRNA from the cells of interest using the methods described in Chapter 6. For optimal results, you will need about 10 μg of poly(A)$^+$ RNA to synthesize enough double-stranded cDNA for a library. However, the reactions will work (albeit less efficiently) if less template is available. Before proceeding, the integrity of the poly(A)$^+$ RNA should be checked by gel electrophoresis (see Chapter 6), using as markers 18S and 28S ribosomal RNAs and purified 9S globin mRNA. As visualized by ethidium-bromide staining of gels or methylene-blue staining of nitrocellulose filters, the poly(A)$^+$ RNA should form a continuous smear (~ 10S - ~ 30S) with most of the molecules migrating at about 16S–18S. There is usually ribosomal RNA present in the poly(A)$^+$ RNA even after two cycles of selection on oligo(dT) columns. The sharpness of the ribosomal RNA bands provides a rough indication of whether the mRNA is degraded.

2. Prepare sterile stock buffers and solutions for first-strand synthesis:

   1 M Tris · Cl (pH 8.3) at 42° (the pH should be measured at 42°C since the pH of Tris changes with temperature)
   1 M KCl
   250 mM $MgCl_2$
   700 mM β-mercaptoethanol (add 50 μl of a concentrated [14 M] solution to 950 μl of H_2O
   dNTP solution (containing all four dNTPs [20 mM] in 0.01 M Tris · Cl [pH 8.0])
   oligo(dT)$_{12-18}$ primer (1 mg/ml in H_2O)
   100 mM methylmercuric hydroxide (see Chapter 6 for preparation and precautions to be taken in handling)

3. Estimate the volume of the AMV reverse transcriptase required. For 10 μg of poly(A)$^+$ mRNA, you will need approximately 40 units of reverse transcriptase. Most enzyme preparations contain 5–20 units/μl.

   The contribution of the storage buffer to the final composition of the reaction mixture must be considered. For example, the standard reverse transcriptase storage buffer contains 200 mM potassium phosphate (pH 7.2). Therefore, to obtain the optimum monovalent cation concentration (140–150 mM K^+), the amount of stock potassium chloride solution included in the reaction mixture must be reduced appropriately if large volumes of reverse transcriptase are added. Moreover, to prevent the added phosphate buffer (pH 7.2) from lowering the final pH of the reac-

tion (optimally, pH 8.3), a relatively high concentration of Tris (100 mM) is used.

For prolonged storage, reverse transcriptase should be kept in small aliquots at −70°C. The enzyme subunits dissociate with time at −20°C. Therefore, only the working solution should be kept at −20°C.

4. Because many batches of reverse transcriptase are contaminated with RNase, potent inhibitors of RNase (RNasin or vanadyl-ribonucleoside complexes) are routinely included in the reaction. Although both types of inhibitor are effective, RNasin has a slight advantage in that it is readily removed by a single extraction with phenol/chloroform.

   RNasin should be used at a final concentration of 0.5 units/μl of reaction mixture.

   Vanadyl-ribonucleoside complexes are prepared as follows. Thaw the stock solution (200 mM; see page 188) immediately before use, centrifuge for 2 minutes (in an Eppendorf centrifuge), and dilute to 10 mM with water. The final concentration in the reaction mix is 1 mM; higher concentrations inhibit reverse transcriptase.

5. In an autoclaved Eppendorf tube, dry down approximately 50 pmoles (~ 40 μl) of each of the four [α-^{32}P]dNTPs (sp. act. = 800 Ci/mM; supplied in ethanol/water [50% v/v]).

   In this case, [α-^{32}P]dNTPs supplied in ethanol have some advantage over those supplied as stabilized aqueous solutions. The latter, which contain Tricene buffer at pH 6.0, would occupy about a third of the reaction volume and would change the pH of the reaction.

   If only aqueous [α-^{32}P]dNTPs are available, the following changes should be made to the reaction mixture:

   a. Make up a solution that contains three unlabeled dNTPs at a concentration of 20 mM and one unlabeled dNTP at a concentration of 10 mM. Use 2.5 μl of this composite solution per 50 μl of reaction mixture.

   b. Add to the reaction 10 μl (100 μCi) of the [α-^{32}P]dNTP present in the composite solution at low concentration.

6. Set up the reaction mixture. A reasonable reaction volume is 50 μl. Smaller volumes are more difficult to handle, and the presence of impurities in the radioactive triphosphates (especially after storage for more than one half-life) may lead to the inhibition of the reaction.

   A larger reaction volume requires more [α-^{32}P]dNTP to achieve the same amount of incorporation into DNA and is unnecessarily expensive.

   a. To the dried down radioactive triphosphates add:

1 mg/ml mRNA	10 μl (10 μg)
100 mM methylmercuric hydroxide	1 μl

Let stand at room temperature for 10 minutes. This treatment denatures the RNA and increases the yield of full-length cDNA from some mRNA templates.

b. Add 2 μl of 700 mM β-mercaptoethanol and 5 μl of 10 mM vanadyl-ribonucleoside complexes (or 2 μl of RNasin, 25 units). Let stand at room temperature for 5 minutes. The β-mercaptoethanol, which is necessary for the stability of reverse transcriptase, is added at this point to sequester the mercury ions since these ions otherwise would inhibit reverse transcription.

c. Add:

1 mg/ml oligo$(dT)_{12-18}$	10 μl (10 μg)
1 M Tris · Cl (pH 8.3)	5 μl
1 M KCl	7 μl
250 mM $MgCl_2$	2 μl
20 mM dNTPs	2.5 μl
H_2O to a final volume of	50 μl

d. Add 2 μl (40 units) of reverse transcriptase, mix well by vortexing, and centrifuge briefly in an Eppendorf centrifuge to eliminate the bubbles that are generated by the presence of Triton X-100 in the enzyme storage buffer. Incubate at 42°C for 1-3 hours.

The final reaction conditions for first-strand synthesis are:

100 mM Tris · Cl (pH 8.3)
10 mM $MgCl_2$
140 mM KCl
100 μg/ml oligo$(dT)_{12-18}$
2 mM methylmercuric hydroxide
20 mM β-mercaptoethanol
1 mM vanadyl-ribonucleoside complexes, *or*
0.5 units/μl RNasin
1 mM each dNTPs
100 μg/ml poly$(A)^+$ RNA
400-800 units/ml reverse transcriptase

7. Stop the reaction by adding 2 μl of 0.5 M EDTA (pH 8.0), followed by 25 μl of 150 mM NaOH.

Note. It is important that the concentration of EDTA is sufficiently high to chelate all the divalent magnesium ions; otherwise an insoluble magnesium hydroxide-DNA complex will form when the sodium hydroxide is added.

8. Incubate for 1 hour at 65° or for 8 hours at 37°C to hydrolyze the mRNA template.

9. Neutralize the solution by adding

1.0 M Tris · Cl (pH 8.0)	25 μl
1.0 N HCl	25 μl

10. Measure the total amount of radioactivity in the reaction and the amount of material incorporated into TCA-precipitable material, as described on page 473.

11. Calculate the yield of cDNA from the percent of dNTPs incorporated. In theory, it is possible to synthesize an amount of cDNA equal in weight to the RNA template. In practice, the yield of the first-strand reverse transcriptase reaction is usually no more than 10–30% of the weight of poly(A)$^+$ RNA added.

12. Extract the remainder of the reaction with an equal volume of phenol/chloroform. After centrifugation, transfer the aqueous phase to a fresh Eppendorf tube. Reextract the organic phase with an equal volume of 10 mM Tris · Cl (pH 8.0), 100 mM NaCl, and 1 mM EDTA. Combine the two aqueous phases.

13. Separate the cDNA from unincorporated dNTPs and the products of alkaline hydrolysis of the template by chromatography on Sephadex G-100 as follows. Layer the combined aqueous phases on a column (~2 ml bed volume) of Sephadex G-100. Collect 0.2-ml fractions and measure the amount of radioactivity by Cerenkov counting. Pool the fractions in the excluded volume that contain radioactivity (see Fig. 7.4). Remove an aliquot (20,000 cpm) of the cDNA for analysis by gel electrophoresis. Precipitate the remainder of the cDNA with ethanol. Alternatively, the unincorporated dNTPs can be removed by spun-column chromatography (see page 466).

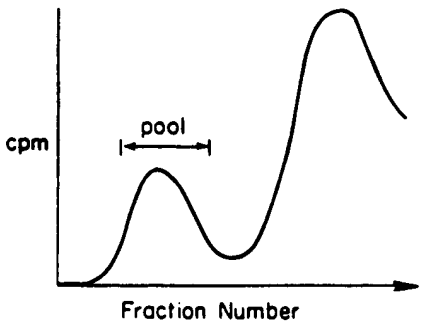


Figure 7.4

14. Measure the size of the first-strand cDNA by electrophoresis through a 1.4% alkaline agarose gel (see page 171).

    Apply 20,000 cpm of the cDNA to the gel. For molecular-weight markers, use a mixture of end-labeled restriction endonuclease fragments of pBR322 DNA (see page 115). Apply 3000 cpm of the markers to the gel. Continue electrophoresis until the bromocresol green has migrated half the length of the gel.

    Fix the DNA by immersing the gel for 30 minutes in each of two changes of 7% trichloroacetic acid.

    Wash the gel briefly in water and blot off any excess fluid. Cover with Saran Wrap and expose for autoradiography (Kodak XR film or equivalent) at room temperature for several hours without intensifying screens.

    If synthesis of the first strand was successful, a smear of radioactivity will be seen (from 100 nucleotides to the size of the largest species in the RNA preparation). Unless the poly(A)$^+$ RNA has been prepared from a differentiated cell type that contains one or more species of highly abundant mRNA, specific bands will not be detected.

Second-strand Synthesis

1. Recover the first-strand cDNA by centrifugation (10 minutes at 4°C in an Eppendorf centrifuge).

2. Resuspend the cDNA in 50 μl of H_2O. Add 50 μl of 2× second-strand buffer. Set 4.0 μl aside for later analysis by nuclease S1 and gel electrophoresis.

   2× Second-strand buffer

   0.2 M HEPES (pH 6.9)
   20 mM $MgCl_2$
   5 mM dithiothreitol
   0.14 M KCl
   1 mM of each of the four dNTPs

3. Add 20–50 units of the Klenow fragment of *E. coli* DNA polymerase I for every microgram of first-strand cDNA in the reaction. The volume of enzyme added should not exceed 15% of the total volume of the reaction, otherwise synthesis of the second cDNA strand may be inhibited by glycerol and phosphate in the enzyme storage buffer. The concentration of enzyme in many commercial preparations is quite low (1–2 units/μl), and it is often necessary to increase the volume of the second-strand reaction in order to accommodate the amount of enzyme required.

   Incubate at 15°C for 20 hours. The long incubation period allows the enzyme to find first-strand cDNA molecules with hairpin loops at their 3′ ends. Presumably, these structures are quite unstable and transient, and the enzyme must wait for and catch molecules in this unlikely configuration in order to begin synthesis of the second DNA strand. Some workers denature the first strand of cDNA by treatment with methylmercuric hydroxide before beginning second-strand synthesis. We have not found this procedure to make any difference to either the efficiency of the second-strand synthesis or to the size of the double-stranded cDNA product.

4. Stop the reaction by adding 2.0 μl of 0.5 M EDTA.

5. Remove two 2.0-μl aliquots from the reaction mixture for later analysis by gel electrophoresis. Extract the remainder of the sample with an equal volume of phenol/chloroform. Separate the double-stranded cDNA from unincorporated dNTPs by chromatography on Sephadex G-50, as described on pages 464–467. Precipitate the DNA with ethanol.

6. Even if the length-distribution of the population of double-stranded cDNA appears to be correct, it is highly unlikely that second-strand synthesis has been completed in all of the molecules. Truncated double-stranded cDNA is thought to arise because of the presence in the template strand of sequences that cause the Klenow polymerase to pause or stop (strong-stop sequences). Because such stopping points are different for Klenow polymerase and reverse transcriptase, it is possible to obtain a greater yield of full-length, double-stranded cDNA by carrying out a reaction with reverse transcriptase after the reaction with the Klenow enzyme has been completed.

   Dissolve the cDNA in 20 μl of H_2O. Add:

1 M Tris · Cl (pH 8.3)	5	μl
1 M KCl	7	μl
250 mM $MgCl_2$	2	μl
a solution containing all four dNTPs at a concentration of 20 mM	2.5	μl
700 mM β-mercaptoethanol	2	μl
H_2O	to 48	μl
reverse transcriptase	2	μl (40 units)

   Incubate at 42°C for 1 hour.

8. Stop the reaction by adding 2.0 μl of 0.5 M EDTA. Remove two 1-μl aliquots for analysis by gel electrophoresis. Extract the remainder of the sample with an equal volume of phenol/chloroform. Separate the double-stranded cDNA from unincorporated dNTPs by chromatography on Sephadex G-50 as described on pages 464-467. Precipitate the double-stranded cDNA with ethanol.

9. Apply samples from steps 2 (2 μl of the 4-μl aliquot), 5, and 8 to an alkaline, 1.4% agarose gel. Run the gel and locate the position of the cDNA by autoradiography, as described on page 171. For molecular-weight markers, use a set of end-labeled fragments of pBR322 DNA (see pages 115ff).

   If synthesis of the second strand is successful, the length of the double-stranded cDNA calculated from its rate of migration through the alkaline gel should be approximately twice that of the first strand. This is because the first and second strands are covalently joined by the hairpin loop.

DIGESTION WITH NUCLEASE S1

The amount of nuclease S1 required is determined by carrying out a set of pilot-scale reactions, each containing approximately 2000 cpm of ^{32}P-labeled, double-stranded cDNA.

1. Dissolve double-stranded cDNA in 50 μl of a solution of 1 mM Tris · Cl (pH 7.6) and 0.1 mM EDTA. Measure the amount of radioactivity by Cerenkov counting.

2. From the solution, remove an aliquot containing 10,000 cpm of ^{32}P. Add 10 μl of 10× nuclease-S1 buffer and sufficient water to bring the volume to 100 μl. Freeze the remainder of the double-stranded cDNA.
   Dispense 20-μl aliquots in five Eppendorf tubes. To each aliquot, add 0, 1, 2, 4, or 6 units of nuclease S1. Incubate at 37°C for 30 minutes.

   10× Nuclease-S1 buffer

   2 M NaCl
   0.5 M sodium acetate (pH 4.5)
   10 mM $ZnSO_4$
   5% glycerol

3. Add 1 μl of 0.5 M EDTA to stop the reactions. Analyze each sample on a 1.4% alkaline gel, using a set of end-labeled fragments of pBR322 DNA as molecular-weight markers (see page 115). Be sure to include on the gel a sample of the first-strand cDNA that was set aside for this purpose.
   Locate the position of the DNA by autoradiography. To obtain the result as quickly as possible:

   a. Fix the gel in 7% trichloroacetic acid (30 minutes, 2 changes).
   b. Wash the gel briefly with water.
   c. Dry down the gel onto Whatman 3MM paper.

   Alternatively:

   a. Soak the gel for 45 minutes in 0.5 M Tris · Cl (pH 7.5) and 1.0 M NaCl.
   b. Transfer the DNA to a nitrocellulose filter by Southern blotting (see pages 382ff).

   Expose the dried-down gel or nitrocellulose filter for autoradiography using Kodak XR film, or its equivalent, with intensifying screens. An overnight exposure at −70°C should be sufficient. Digestion with increasing amounts of nuclease S1 should yield populations of molecules of decreasing modal size. Choose the concentration of enzyme that yields molecules whose modal distribution is the same as that of the first-strand cDNA.

Note. S. Zeitlin and A. Efstratiadis (pers. comm.) have suggested an alternative procedure for calibrating the nuclease-S1 reaction. The procedure is based on the observation that the plasmid PML-21 (Hershfield et al. 1974) contains an inverted-repeat sequence of 1050 bp that is part of the kanamycin-resistance element Tn*903* (Sim et al. 1979). The assay is carried out by linearizing the plasmid DNA by digestion with a restriction enzyme, denaturing by boiling and quick cooling, digesting with nuclease S1, and analyzing the product on native and denaturing agarose gels. A successful digest is indicated by the presence of a discrete band of 1050 bp in both the native and denaturing gels.

a. Digest 10 μg of PML-21 DNA with *Eco*RI in 100 μl of *Eco*RI buffer. Note that *Eco*RI does not cleave within the inverted-repeat sequence (1 μg plasmid DNA ≡ 100 ng of hairpin DNA).

b. Add 900 μl of H_2O, mix, and divide the sample into 100-μl aliquots.

c. Denature the DNA by boiling and then cool the samples rapidly by plunging them into a dry-ice/ethanol bath.

d. Allow the DNA solutions to thaw in ice and then to each tube add 100 μl of ice-cold 2× nuclease-S1 buffer containing 0, 10, 20, 40, 60, 80, or 100 units of nuclease S1. Incubate at 37°C for 30 minutes. Usually, approximately 50 units of nuclease S1 are required to digest 1 μg of denatured PML-21 DNA.

e. Analyze 25 μl of each sample by electrophoresis through neutral and denaturing 2.0% agarose gels. Visualize the DNA by staining with ethidium bromide.

4. Digest the remainder of the double-stranded cDNA with the appropriate amount of nuclease S1.

5. Stop the reaction by addition of 2 μl of 0.5 M EDTA. Add 2 M Tris base to a final concentration of 0.05 M. Extract the solution once with phenol/chloroform. Precipitate the DNA with ethanol.

6. Redissolve the cDNA in 18 μl of TE (pH 8.0). Add 2 μl of 3 M NaCl. Fractionate the cDNA into size classes by passage through a 1-ml column of Sepharose CL-4B (see pages 464–465) equilibrated in 10 mM Tris · Cl (pH 8.0), 0.3 M NaCl, and 1 mM EDTA. Collect 50-μl fractions. Assay an aliquot of each fraction by electrophoresis through an alkaline agarose gel (1.4%). For molecular-weight markers, use a set of end-labeled fragments of pBR322 DNA. Locate the position of the cDNA by autoradiography as described in step 2 above. Pool the fractions that contain cDNA molecules greater than 500 bp in length. Precipitate the cDNA with ethanol.

CLONING DOUBLE-STRANDED cDNA

Homopolymeric Tailing of Vector DNA with Poly(dG)

1. Digest 55 μg of vector DNA (pBR322, pAT153, or pXf3) with *Pst*I. Check that the digestion is complete by analyzing a small sample by electrophoresis through a 1% agarose minigel.

2. Purify the linear DNA by electrophoresis through a preparative 1% agarose gel. (A slot 4 cm long and 4 mm deep will be required to avoid overloading).

3. Extract the DNA from the gel by electroelution, as described on pages 164ff. This purification step is important for two reasons. First, it removes any RNA or low-molecular-weight DNA contaminating the plasmid DNA or the restriction enzyme. Second, it separates the linear, plasmid DNA from any circular molecules that have not been digested with *Pst*I. Such molecules contribute significantly to the background of nonrecombinant transformants when the vector DNA preparation is used to transform *E. coli*.

4. Dissolve the linear plasmid DNA in 55 μl of H_2O. Add 55 μl of 2× tailing buffer.

   2× Tailing buffer

   0.4 M potassium cacodylate
   50 mM Tris · Cl (pH 6.9)
   4 mM dithiothreitol
   1 mM $CoCl_2$
   2 mM [^{3}H]dGTP (sp. act. = 12 Ci/mmole)
   500 μg/ml bovine serum albumin

   (The potassium cacodylate should be diluted from a 1 M solution that has been passed through a Chelex column equilibrated with potassium ions.)

5. Transfer 10 μl to a fresh tube and incubate for 10 minutes at 37°C. Store the remainder of the sample at −20°C.

6. Add 2 units of terminal transferase to the 10-μl aliquot. Mix and continue incubation at 37°C.

7. Remove 1-μl aliquots after 0, 1, 2, 5, 10, 20, 30, and 40 minutes of incubation. Spot the aliquots onto DE-81 filter discs. Wash the discs and count the radioactivity as described on page 473. Calculate how many dG residues have been added per end.

8. Incubate the remaining 100 μl of linear plasmid DNA for 10 minutes at 37°C. Add 20 units of terminal transferase and incubate for the time that results in the addition of 15–20 dG residues per end.

9. Stop the reaction by chilling to 0°C. Add 10 μl of 0.5 M EDTA (pH 8.0). Extract once with phenol/chloroform.

10. Separate the homopolymerically tailed DNA from low-molecular-weight contaminants by chromatography on a column of Sephadex G-100 equilibrated in 1× annealing buffer. Store the tailed DNA in aliquots at −20°C.

    Note. Sephadex G-100 cannot be used in spun columns because the centrifugation crushes the beads.

    10× Annealing buffer

    1 M NaCl
    0.1 M Tris · Cl (pH 7.8)
    1 mM EDTA

11. To check that the tailed vector is functional and is not contaminated by uncut or unit-length, untailed plasmid DNA, a trial annealing and transformation of *E. coli* should be carried out:

    a. Add 20–30 dC residues to a small (200–500 bp) fragment of DNA using the procedure described above for dG tailing.

    b. Set up the following annealing reactions:

    Tube A:

uncut plasmid DNA	0.1 μg
H_2O	to 18 μl
10× annealing buffer	2 μl

    Tube B:

dG-tailed vector	0.1 μg
H_2O	to 18 μl
10× annealing buffer	2 μl

    Tube C:

dG-tailed vector	0.1 μg
dC-tailed insert	0.01 μg
H_2O	to 18 μl
10× annealing buffer	2 μl

    c. Heat to 65°C for 5 minutes and allow the DNAs to reanneal by incubating at 57°C for 1–2 hours. Transform *E. coli* strain RR1 (see Chapter 8). The efficiency of transformation by dG-tailed vector alone should be reduced at least 100-fold compared with circular plasmid.

    The efficiency of transformation by the recombination plasmid (tube C) should be at least 10-fold greater than that of dG-tailed vector alone.

Homopolymeric Tailing of Double-stranded cDNA with Poly(dC)

1. Calculate the quantity of double-stranded cDNA synthesized from the amount of [α-^{32}P]dCTP incorporated during first-strand synthesis. Estimate the total number of molecules synthesized from the size distribution of the double-stranded cDNA. Because an accurate measurement of size is not usually possible and the amount of double-stranded cDNA is limited, the rate of addition of homopolymeric dC tails is tested using 5 μg of plasmid DNA linearized with *Pst*I. This is not as irrational as it sounds. Terminal transferase reactions are carried out with the enzyme in vast excess, so that the number of residues added is essentially independent of DNA concentration.

   Set up a series of pilot reactions with terminal transferase by using 0.5 μg of linearized plasmid DNA and 2 units of terminal transferase, exactly as described on page 238 except that [^{3}H]dCTP is used instead of dGTP. Take samples after 30 seconds, 1 minute, 2 minutes, and 4 minutes of incubation. Spot the aliquots onto DE-81 filter discs. Wash the discs and count the radioactivity as described on page 473. Calculate how many dC residues have been added per end.

2. Recover the double-stranded cDNA by centrifugation for 10 minutes at 4°C in an Eppendorf centrifuge. Dry the DNA pellet briefly under vacuum.

3. Dissolve the double-stranded cDNA in 25 μl of H_2O. Add 25 μl of 2 × tailing buffer prepared with [^{3}H]dCTP (see page 239). Incubate for 10 minutes at 37°C.

4. Add 5 units of terminal transferase for every microgram of double-stranded cDNA in the reaction. Incubate for the time calculated from the pilot reactions to allow addition of 15–20 dC residues.

5. Stop the reaction by chilling to 0°C. Add 10 μl of 0.5 M EDTA (pH 8.0). Extract once with phenol/chloroform.

6. Separate the tailed DNA from low-molecular-weight contaminants by chromatography through a column of Sephadex G-100 equilibrated in annealing buffer or by spun-column chromatography using Sephadex G-50 (see page 466). Store the tailed DNA at −20°C.

ANNEALING VECTOR AND DOUBLE-STRANDED cDNA

1. Mix equimolar amounts of dC-tailed cDNA and dG-tailed vector in annealing buffer at a final concentration of 1 ng/μl.

   Annealing buffer

   0.1 M NaCl
   10 mM Tris · Cl (pH 7.8)
   1.0 mM EDTA

2. Heat to 65°C for 5 minutes and allow the DNAs to anneal by incubating at 57°C for 1-2 hours. Store the reannealed DNAs at −20°C.

3. Carry out the transformation of *E. coli* strain RR1 by using the protocol described on page 254.

CLONING DOUBLE-STRANDED cDNA BY SEQUENTIAL ADDITION OF LINKERS

1. Synthesize the first and second strands of cDNA, as described previously. Do not digest the double-stranded hairpin DNA with nuclease S1.

2. Prepare two sets of kinased linkers, as described on page 396.

3. To maximize the number of molecules with a perfectly blunt end, the hairpin double-stranded cDNA is treated with the Klenow fragment of *E. coli* DNA polymerase I in the presence of all four dNTPs.

   Dissolve approximately 2 μg of double-stranded cDNA in 11 μl of TE (pH 7.4). Add:

10× repair buffer	2 μl
1.0 mM dATP	1.25 μl
1.0 mM dCTP	1.25 μl
1.0 mM dGTP	1.25 μl
1.0 mM dTTP	1.25 μl
Klenow fragment of DNA polymerase I	1 unit (~ 1 μl)

   Incubate for 30 minutes at room temperature.

   10× Repair buffer

   0.5 M Tris · Cl (pH 7.4)
   70 mM $MgCl_2$
   10 mM dithiothreitol

4. The first linker is added to the blunt end of the hairpin double-stranded cDNA, i.e., at the end corresponding to the 3′ terminus of the original mRNA. Enough kinased linkers should be added to achieve a 1:1 mass ratio with double-stranded cDNA.

   At the end of the repair reaction (step 3), add 30 μl of 2× blunt-end ligation buffer. Then add:

2 μg kinased linkers in a volume of	4 μl
10 Weiss units T4 polynucleotide ligase	~ 1 μl
20 units RNA ligase	2 μl

   Incubate for 12–16 hours at 4°C.

2× Blunt-end ligation buffer

50 mM Tris · Cl (pH 7.4)
10 mM $MgCl_2$
10 mM dithiothreitol
0.5 mM spermidine
2 mM ATP
2.5 mM hexamine cobalt chloride
20 μg/ml BSA

Blunt-end ligation buffer should be stored in small aliquots at −20°C.

5. Stop the reaction by addition of 2 μl of 0.5 M EDTA. Extract once with phenol/chloroform. Precipitate the DNA with ethanol.

6. Dissolve the double-stranded cDNA in 45 μl of a solution of 1 mM Tris · Cl (pH 7.6) and 0.1 mM EDTA. Cleave the hairpin loop with nuclease S1, as described on pages 237ff.

7. Stop the reaction by addition of 2 μl of 0.5 M EDTA and 2.5 μl of 1 M Tris base. Extract once with phenol/chloroform.

8. Separate the double-stranded cDNA from low-molecular-weight contaminants by chromatography on Sephadex G-100. Precipitate the double-stranded cDNA with ethanol.

9. Repair the double-stranded cDNA with the Klenow fragment of *E. coli* DNA polymerase I, as described in step 3 (page 243).

10. Add the second kinased linker as described in step 4 (page 243).

11. Dilute the ligation reaction so that the composition of the buffer is suitable for digestion of the linkered DNA by the appropriate restriction enzymes. Add 50 units of each enzyme for every microgram of linker used in the ligation reactions. Incubate for 6–8 hours at the appropriate temperature.

12. Terminate the reaction by addition of EDTA to a final concentration of 10 mM. Extract once with phenol/chloroform. Precipitate the DNA with ethanol.

13. Redissolve the double-stranded cDNA in 10 μl of H_2O. Add 10 μl of

    0.6 M NaCl
    20 mM Tris · Cl (pH 8.0)
    2 mM EDTA

Fractionate the cDNA into size classes by passage through a 1-ml column of Sepharose CL-4B equilibrated in the same buffer. Collect 50-μl fractions. Assay an aliquot of each fraction by electrophoresis through an alkaline agarose gel (1.4%), using as molecular-weight markers a set of end-labeled fragments of pBR322 DNA. Locate the position of the cDNA by autoradiography (see page 470). Pool the fractions that contain cDNA molecules greater than 500 bp in length. Precipitate the cDNA with ethanol.

14. Prepare the vector DNA as follows. Digest 50 μg of plasmid with the appropriate restriction enzymes. Purify the desired fragment of DNA either by gel electrophoresis (see Chapter 5) or by sucrose gradient centrifugation, essentially as described by Kurtz and Nicodemus (1981). The gradient (10–40% [w/v] sucrose in 10 mM Tris · Cl [pH 7.9], 1 mM EDTA, and 1 M NaCl) can be poured in the conventional way in a Beckman SW41 centrifuge tube, or it can be made by three cycles of freezing at −70°C and thawing at 4°C of a 20% (w/v) solution of sucrose in the same buffer.

    Up to 100 μg of DNA may be loaded onto a single gradient, which is centrifuged for 34 hours at 40,000 rpm in an SW41 rotor at 4°C. Fractions (0.4 ml) are collected from the bottom of the tube and 15-μl aliquots are analyzed by electrophoresis on an agarose gel. Fractions containing the vector DNA are pooled, diluted threefold with water to reduce the sucrose concentration, and precipitated with ethanol.

    To check that the vector DNA is functional and is not contaminated by uncut or unit-length, linear plasmid DNA, trial ligation and transformation of *E. coli* are carried out.

    a. Prepare a small DNA fragment (200–500 bp) with ends that are compatible with those of the vector.

    b. Set up the following ligation reactions:

    ligation tube A:

uncut plasmid DNA		0.1 μg
H_2O	to	18 μl
10× ligation buffer		2 μl
ligase		5 Weiss units

Ligation tube B:

vector DNA	0.1 μg
H_2O	to 18 μl
10× ligation buffer	2 μl
ligase	5 Weiss units

Ligation tube C:

vector DNA	0.1 μg
small fragment of DNA	0.01 μg
H_2O	to 18 μl
10× ligation buffer	2 μl
ligase	5 Weiss units

Incubate each sample at 4°C for 12-16 hours.

10× Ligation buffer

0.5 M Tris (pH 7.4)
0.1 M $MgCl_2$
0.1 M dithiothreitol
10 mM spermidine
10 mM ATP
1 mg/ml BSA

c. Transform *E. coli* strain DH1 or HB101 (see Chapter 8) with 10 ng of the ligated DNA.

The efficiency of transformation of *E. coli* by the vector DNA (ligation tube B) should be reduced 10^4-fold compared with undigested plasmid. The efficiency of transformation of *E. coli* by the reconstructed plasmid (ligation tube C) should be at 10-fold to 100-fold greater than that of the vector alone.

15. Mix the appropriate amount of vector with double-stranded cDNA to achieve a molar ratio of vector to cDNA of 5:1. Heat to 68°C for 10 minutes. Chill in ice. Add water and 10× ligase buffer so that the final concentration of vector DNA is 1.5 μg/ml in 1× ligation buffer.

16. Add 10 Weiss units of T4 polynucleotide ligase for every microgram of vector DNA in the reaction. Incubate for 12-16 hours at 12°C.

17. Add EDTA to a final concentration of 10 mM. Extract once with phenol/chloroform and precipitate the DNA with ethanol.

18. Carry out transformation of *E. coli* DH-1 by using one of the protocols given in Chapter 8.